UNISIST

Principes directeurs pour l'établissement et le développement de thesaurus monolingues

SC/WS/555
Paris, septembre 1973

Organisation des Nations Unies pour l'éducation, la science et la culture

ORGANISATION DES NATIONS UNIES
POUR L'EDUCATION, LA SCIENCE ET LA CULTURE

Principes directeurs pour l'établissement et le développement de thesaurus monolingues

(Le présent document constitue la version révisée du document SC/WS/500, Paris, 22 décembre 1971 : Principes directeurs pour l'établissement et le développement de thesaurus monolingues destinés à la recherche documentaire)

SC/WS/555

TABLE DES MATIERES

page

1. INTRODUCTION 5

1.1 Domaine d'application 5
1.2 But 5

2. DEFINITIONS 5

2.1 Définition des thesaurus 5
2.2 Relations avec les autres langages documentaires 7

3. STRUCTURE DES THESAURUS : PRINCIPES ET METHODES 7

3.1 Descripteurs 7
3.2 Règles de forme 9

3.2.1 Mots composés 9
3.2.2 Représentation de notions par plusieurs descripteurs . 9
3.2.3 Forme des mots 11
3.2.4 Forme substantivée 12
3.2.5 Nombre 12
3.2.6 Adjectifs 12
3.2.7 Abréviations et sigles 13
3.2.8 Emploi des caractères 13

3.3 Comment éviter l'ambiguïté 14

3.3.1 Homonymes 14
3.3.2 Notes d'application (ou notes d'usage) 14
3.3.3 Définitions 15
3.3.4 Traduction 15
3.3.5 Indication des sources 15

3.4 Relations entre les descripteurs 16

3.4.1 Généralités 16
3.4.2 Relation d'équivalence (relation préférentielle) . . . 16
3.4.3 Relations hiérarchiques 18
3.4.4 Relation associative (relation d'affinité) 19
3.4.5 Symboles pour références de renvoi 20

4. PRESENTATION ET DISPOSITION DU THESAURUS 21

4.1 Introduction 21
4.2 Partie principale du thesaurus 21
4.3 Parties annexes du thesaurus 22

4.3.1 Généralités 22
4.3.2 Index alphabétiques 23
4.3.3 Listes systématiques 23
4.3.4 Représentations graphiques 24

4.4 Ordre de classement (classement alphabétique) 25

5. METHODE DE CONSTRUCTION DU THESAURUS 26

5.1 Comment éviter les doubles emplois 26
5.2 Notification préalable 26
5.3 Principes généraux 27
5.4 Différentes méthodes 27
5.5 Choix des descripteurs 28

page

5.5.1 Collecte des termes 28
5.5.2 Vérification 29
5.5.3 Evaluation 30
5.5.4 Choix 30
5.6 Procédé d'enregistrement 30
5.7 Essais d'application 31

6. MISE A JOUR 31
6.1 Vérification périodique 32
6.2 Elimination de descripteurs 32
6.3 Choix de nouveaux descripteurs 32

7. REFERENCES 34
7.1 Liste des recommandations de l'ISO liées aux principes directeurs ci-dessus 34
7.2 Sources pour les dictionnaires, glossaires et thesaurus . . 34

HISTORIQUE

Les présents principes directeurs sont le résultat d'un effort conjoint de l'Unesco et de l'ISO. Comme les thesaurus sont des instruments qui jouent un rôle important dans le processus d'information, les deux organisations ont reconnu la nécessité de coopérer pour éviter tout chevauchement dans leurs travaux.

Au cours de la réunion d'ISO/TC 46 (Lisbonne, 1971), il a donc été décidé que l'ISO et l'Unesco rédigeraient en commun un document intitulé "Principes directeurs pour la préparation de thesaurus monolingues" qui servirait, d'une part, d'avant-projet de l'ISO et, d'autre part, de version révisée des Principes directeurs de l'Unesco.

Un sous-groupe de travail conjoint Unesco-ISO/TC 46 a élaboré les premiers "Principes directeurs pour l'établissement et le développement de thesaurus monolingues destinés à la recherche documentaire" en se fondant principalement sur les documents ci-après :

- Unesco, Principes directeurs pour l'établissement et le développement de thesaurus scientifiques et techniques monolingues destinés à la recherche documentaire. Paris, 10 août 1970.
- Avant-projet allemand, Thesaurus Guidelines, décembre 1970, ISO/TC 46 (Germany-2) 984E.
- Premier projet des American National Standard Guidelines for Thesaurus Structure, Construction and Use. ANSI Z39, février 1971.
- Les observations des organismes membres d'ISO/TC 46 et d'ISO/INFCO ont été prises en considération ainsi que les actes de la conférence internationale sur les principes généraux de l'établissement des thesaurus qui s'est réunie à Varsovie en mars 1970. Il a également été tenu compte des conclusions du Groupe de travail ad hoc d'ISO/TC 46 sur l'établissement des thesaurus (Lisbonne, 1971), ainsi que des observations formulées par de nombreux experts et organisations compétentes.

Les "Principes directeurs " mentionnés ci-dessus ont ensuite été largement diffusés en tant que document de l'Unesco SC/WS/500 ("Principes directeurs pour l'établissement et le développement de thesaurus monolingues destinés à la recherche documentaire") ayant le même contenu que le document ISO/DIS 2788.

D'autres observations et suggestions ont été reçues qui ont abouti finalement aux présents "Principes directeurs pour l'établissement et le développement de thesaurus monolingues", élaborés par le sous-groupe de travail conjoint Unesco-ISO/TC 46.

1. INTRODUCTION

1.1 DOMAINE D'APPLICATION

Les présents principes directeurs sont applicables à la préparation et au développement de thesaurus unilingues permettant le stockage et la recherche d'informations pour les besoins de la documentation dans tous les domaines. Cependant, le codage des formules mathématiques ou des structures chimiques n'est pas traité dans le présent document. Ces principes directeurs sont d'application générale. Il y a lieu de les compléter pour tenir compte des particularités de la langue employée.

1.2 BUT

Les centres d'information et de documentation ont besoin de méthodes pratiques pour représenter les notions de façon claire et simple, et pour les ordonner en précisant les relations entre ces notions. L'application de ces méthodes et de ces règles de terminologie permet d'introduire un système rationnel et efficace d'analyse, de classification et de recherche dans les systèmes documentaires.

De nombreux organismes, dans plusieurs pays, ont constitué ou sont en train de constituer des thesaurus pour leurs propres besoins. Il est évident que le transfert des informations rend nécessaire une série normalisée de règles de construction des thesaurus. Ces principes directeurs seront utiles :

- à chaque organisme, pour l'aider à constituer et à mettre au point un thesaurus ;
- pour faciliter le transfert des informations entre les organismes.

Ces principes directeurs sont destinés à faciliter la préparation et la mise au point des thesaurus quel que soit leur mode d'exploitation, manuel ou automatique. Finalement, ces principes directeurs ont été conçus en vue de fournir, actuellement et dans l'avenir, une base assurant la compatibilité de thesaurus établis simultanément dans la plupart des domaines des sciences fondamentales ou appliquées.

2. DEFINITIONS

2.1 DEFINITION DES THESAURUS

Un thesaurus peut être défini soit selon sa fonction soit selon sa structure.

Du point de vue de sa fonction, un thesaurus est un instrument de contrôle de la terminologie utilisé pour traduire en un langage plus strict (langage

documentaire, langage d'information) le langage naturel employé dans les documents et par les indexeurs ou les utilisateurs.

Du point de vue de sa structure, le thesaurus est un vocabulaire, contrôlé et dynamique, de termes ayant entre eux des relations sémantiques et génériques, et qui s'applique à un domaine particulier de la connaissance.

Etant un sous-ensemble structuré du langage naturel, il décrit le contenu analytique des documents, des objets ou des ensembles de données. La terminologie nécessaire pour décrire les caractéristiques du document ou de l'objet, qui comprend par exemple des termes non descriptifs du contenu, comme les noms d'auteurs, les collectivités-auteurs, les revues, musées, emplacements (description bibliographique), n'a pas besoin de faire partie d'un thesaurus mis en forme, bien que ces termes soient généralement nécessaires pour identifier et localiser ce qui est indexé. Ils peuvent, en fait, constituer une annexe du thesaurus.

Un thesaurus spécialisé doit refléter avec précision les informations contenues dans l'ensemble des documents ou autres éléments de la collection à laquelle le thesaurus s'applique. Il doit contenir les termes et les renvois appropriés au sujet, en prenant en considération, à la fois, le langage de la collection de documents et le langage et les besoins des utilisateurs.

Selon la nature des règles de terminologie appliquées, deux types principaux de thesaurus font actuellement l'objet d'études :

a) les thesaurus dont les règles reposent sur un choix de termes préférentiels, c'est-à-dire les thesaurus dans lesquels un seul des termes représentant une notion est admis en vue de l'indexation et de la recherche documentaire ;

b) les thesaurus dont les règles admettent pour l'indexation et la recherche documentaire, tous les termes représentant une notion, mais qui font correspondre à tous ces termes une représentation non ambiguë de la notion (par exemple un code numérique de notion ou une notation)/1. Dans ce type de thesaurus, quand un terme change de sens, il est plus facile de supprimer les synonymes et de modifier les relations antérieures.

---

1. Si les thesaurus à termes préférentiels peuvent être mis à jour manuellement, les thesaurus du deuxième type exigent l'emploi d'une machine pour la mise à jour et la recherche documentaire.

2.2 RELATIONS AVEC LES AUTRES LANGAGES DOCUMENTAIRES

A l'opposé d'un dictionnaire, qui fournit les définitions de mots ou de termes donnés, un thesaurus fournit des mots ou des termes pour exprimer les notions impliquées par les relations entre termes données dans le thesaurus.

Un thesaurus peut être présenté comme un index alphabétique et les termes d'un thesaurus peuvent être utilisés pour en constituer un. Mais le thesaurus lui-même n'est pas un index. Dans l'index d'une collection il doit y avoir une adresse ou un indicateur de localisation pour chaque article de la collection correspondant à chaque terme, alors que le thesaurus ne contient que les termes, sans les adresses ou indicateurs de localisation de l'index. Un thesaurus classe les termes dans des catégories hiérarchiques. En tant que "système de classification de termes", un thesaurus est assez proche des classifications par matières, par exemple la Classification décimale universelle.

Mais tandis que les systèmes hiérarchisés de classification par matières visent à mettre en évidence l'ensemble des relations hiérarchiques, les thesaurus donnent les relations nécessaires à l'indexation et à la recherche documentaire, d'après l'ensemble des documents et les besoins en information des utilisateurs.

Un thesaurus est une sorte de liste normative, c'est-à-dire que les termes préconisés dans un thesaurus particulier sont ceux qui doivent être utilisés pour l'indexation et la recherche documentaire dans un système d'information et de documentation donné. Il existe d'autres catégories de listes normatives, basées sur le langage naturel, telles que les listes de vedettes-matière. Elles ne présentent pas, en général, la structure hiérarchisée d'un thesaurus.

3. STRUCTURE DES THESAURUS

3.1 DESCRIPTEURS

La configuration de chacun des articles et leur classement en relation les uns avec les autres constituent la structure d'un thesaurus. Les renvois, dans un thesaurus, rendent explicite la nature des relations qui lient les articles les uns aux autres dans un réseau de notions.

Les termes retenus dans un thesaurus pour l'indexation sont appelés les descripteurs. Dans les thesaurus à termes préférentiels, les descripteurs sont ces termes préférentiels. Un descripteur peut être défini comme un terme ou un symbole formalisé et homologué dans un thesaurus, et employé pour représenter sans ambiguïté les notions contenues dans les documents et dans les demandes de recherche documentaire.

En vue d'une coopération internationale, deux niveaux peuvent être distingués parmi les descripteurs :

- les descripteurs les plus généraux, qui pourraient faire l'objet d'accords multinationaux après leur traduction en plusieurs langues ;
- les descripteurs plus spécifiques, qui pourraient faire l'objet d'accords dans plusieurs domaines spécialisés.

Peuvent être considérés comme descripteurs :

a) les termes représentant des notions ou des combinaisons de notions ;

b) les termes représentant des entités individuelles. Ces termes sont appelés également noms propres et assimilés (ou identificateurs). Ces noms propres et assimilés peuvent être :

- des noms de projets ;
- des nomenclatures ;
- des numéros ou des symboles d'identification ;
- des noms géographiques ou géopolitiques ;
- des marques de fabrique ;
- des noms de personnes ou d'organisations ;
- des abréviations ou des sigles ;
- d'autres noms propres et assimilés (par exemple de programmes de calculateurs).

Il est conseillé d'utiliser les noms propres et assimilés de la même façon que les autres descripteurs, c'est-à-dire de les relier entre eux. On fera de même pour intégrer dans un thesaurus les nomenclatures admises sur le plan international.

Tous les termes compris dans les thesaurus n'utilisant pas de termes préférentiels peuvent être, en principe, des descripteurs.

Les termes représentant des notions non retenues pour l'indexation doivent être considérés comme des termes non homologués. Ils sont appelés "non-descripteurs".

Dans la plupart des cas, il est utile de donner la possibilité d'une distinction formelle entre les descripteurs et les non-descripteurs. Cela peut être réalisé par l'utilisation des caractères d'imprimerie spéciaux pour les descripteurs ou de signes spéciaux au début et à la fin de chaque descripteur.

3.2 REGLES DE FORME

3.2.1 MOTS COMPOSES

Un descripteur peut comprendre un ou plusieurs mots. En règle générale, le descripteur doit refléter la terminologie du sujet quel que soit le nombre des mots nécessaires pour représenter la notion. Mais dans la mesure du possible, un descripteur doit comprendre le minimum de mots, et de préférence, un seul. Il ne faut pas oublier, cependant, qu'un terme peut perdre de sa clarté à la suite d'une abréviation. Les descripteurs composés sont enregistrés de préférence dans l'ordre naturel des mots (par exemple "génie électrique"), c'est-à-dire sans inversion artificielle de leur ordre. Il peut être utile d'inclure les formes inversées comme non-descripteurs, liés aux descripteurs par des relations préférentielles.

Lorsque la forme inversée est choisie comme tête d'article, c'est-à-dire comme descripteur, il est nécessaire de faire figurer comme synonyme la forme non inversée.

3.2.2 REPRESENTATION DE NOTIONS PAR PLUSIEURS DESCRIPTEURS

Pour limiter le nombre de descripteurs, il est parfois utile de représenter les notions ou les combinaisons de notions par une association de descripteurs.

3.2.2.1 Division des notions ou des combinaisons de mots

En règle générale, il existe deux possibilités :

a) Division (lexicologique) morphologique

Si une notion est représentée par un mot composé ou un ensemble de plusieurs mots, il est souvent possible de diviser cet ensemble en ses composants. Lorsque les mots ainsi obtenus sont associés, ils ne représentent cependant la notion originale que si la division morphologique coïncide avec la division sémantique. C'est dans ce cas seulement que la division morphologique donne de bons résultats.

b) Division sémantique

Dans la division sémantique, une notion est divisée en éléments qui, lorsqu'ils sont combinés, représentent la notion originale. Les descripteurs utilisés ne sont pas nécessairement contenus dans les termes.

Exemple de (a)

PSYCHOLOGIE DES ANIMAUX = PSYCHOLOGIE • ANIMAUX

Exemples de (b)

TOILE-EMERI = TISSU + ABRASIF

GARDE-BOUE = PROTECTION + BOUE

3.2.2.2. Combinaisons de mots ou notions

Les termes peuvent être combinés (ou précoordonnés) avant l'introduction dans le système et doivent donc être recherchés tels quels, ou combinés pendant la recherche (postcoordonnés) pour représenter la notion recherchée. Le choix entre la pré ou la postcoordination relève des considérations suivantes :

a) Postcoordination

La postcoordination doit être limitée aux cas où les descripteurs simples ne sont pas utilisés seuls trop fréquemment et/ou le descripteur précombiné est employé très rarement. Il importe de s'assurer que la combinaison de descripteurs simples représente réellement et sans ambiguïté la même notion que le descripteur précombiné.

Exemple négatif :

TRANSFORMATEUR + HUILE = TRANSFORMATEUR A HUILE
ou HUILE DE TRANSFORMATEUR

b) Précoordination

Les descripteurs précombinés doivent toujours être employés quand :

- les descripteurs simples ont une signification différente selon qu'ils sont employés seuls et/ou qu'ils sont inclus dans un descripteur précombiné.

  Exemple : ACIDE MINERAL = ACIDE NON ORGANIQUE ; tandis qu'un minéral est un produit extrait du sol.

- les descripteurs simples sont utilisés dans d'autres liaisons hiérarchiques que les descripteurs précombinés.

  Exemple : BRISE-VENT = ELEMENT DE CONSTRUCTION QUI ARRETE LE VENT ; tandis qu'une brise est un vent faible.

- le descripteur précombiné est un nom propre, ou bien il est douteux que l'association de descripteurs simples reflète exactement et exclusivement le contenu de la notion.

Dans la plupart des cas il est recommandé d'utiliser des termes génériques comme éléments de précombinaison ; si une notion est représentée par une combinaison de descripteurs simples, cela doit être indiqué par la référence "EM" (USE). Les descripteurs simples, tels qu'ils sont

spécifiés, portent une référence de renvoi "EP" (UF) (par exemple "EPC" (UFC) au non-descripteur précombiné.

Exemple :	RIDEAU D'AIR	EM	RIDEAU + AIR
	RIDEAU	EPC	RIDEAU D'AIR

La combinaison de descripteurs simples doit figurer dans les parties systématiques du thesaurus, et le descripteur précombiné non utilisé, dans les parties alphabétiques comme non-descripteur.

3.2.3 FORME DES MOTS

Une fois qu'on a décidé de retenir un terme donné dans le thesaurus, il faut veiller à ce que ce terme, accompagné des relations représentées, exprime aussi exactement que possible le sens voulu.

a) Orthographe

C'est l'orthographe la plus largement admise qu'il convient d'adopter. Si, par suite de l'existence de plusieurs usages, on est arrivé à accepter plus d'une orthographe, les différentes graphies doivent être introduites dans le thesaurus avec des renvois de l'une à l'autre.

Exemple : CLEF
CLÉ

On peut encore, lorsqu'une telle difficulté surgit, prendre pour arbitre un dictionnaire qui fait autorité.

b) Traduction

De nombreux termes techniques utilisés aujourd'hui sont traduits d'autres langues mais, parfois, un terme latin ou grec ou un terme appartenant à une langue étrangère vivante est incorporé tel quel à un vocabulaire spécialisé se rapportant à une question particulière. Quand le terme étranger et sa traduction putative coexistent, il faut les faire figurer tous les deux avec des renvois de l'un à l'autre.

c) Translittération

Le problème devient plus compliqué encore lorsque la langue étrangère en question s'écrit avec un alphabet différent. La remarque vaut notamment pour les identificateurs. Les normes de translittération de l'ISO doivent être appliquées le plus souvent possible. Lorsqu'on a le choix, il faut donner la préférence au mode de translittération qui n'impose pas de recourir à des signes diacritiques.

Exemple : SATELLITE
SPOUTNIK

3.2.4 FORME SUBSTANTIVÉE

Le descripteur doit être présenté, de préférence, sous forme de substantif ou d'expression équivalant au substantif.

Exemples : "DEMOCRATIE" au lieu de "DEMOCRATIQUE"
"EXECUTION" au lieu de "EXECUTER"

3.2.5 NOMBRE

L'utilisation de descripteurs au singulier ou au pluriel, doit être décidée suivant les usages de la langue du thesaurus. Pour cette décision, il est nécessaire d'établir et de se conformer aux normes nationales. A défaut de normes nationales, il faut suivre les habitudes lexicographiques de la langue. Il arrive qu'un même mot représente des notions différentes, selon qu'il est au singulier ou au pluriel ; en pareil cas, il faut indiquer ces deux formes.

Exemples :		
	MEMOIRE	CISEAU
	MEMOIRES	CISEAUX

En anglais il y a lieu, en général, de mettre les descripteurs au pluriel, particulièrement lorsqu'il s'agit de termes génériques (c'est-à-dire quand les descripteurs représentent des classes d'objets). On présente, par contre, au singulier, les termes indiquant des matériaux, des propriétés (attributs), des procédés, des noms propres et des disciplines.

Exemples :

procédés	propriétés	classes d'objets
ACIDIFICATION	CONDUCTIVITY	TEETH
CALENDERING	OPACITY	STARS
CURING	TEXTURE	PAINTINGS

3.2.6 ADJECTIFS

Il existe, évidemment, un certain nombre de cas où seuls les adjectifs ou d'autres formes non substantivées peuvent être utilisés.

Exemples : SOCIAL
INTERNATIONAL

Une faible proportion de termes consistant en un seul mot sous forme d'adjectif peut être utilisée comme modificateurs (continu, horizontal). Puisque les adjectifs peuvent être associés à des noms et enregistrés comme descripteurs composés (précoordonnés), la décision d'enregistrer des adjectifs seuls doit être dictée par un souci de souplesse et d'efficacité. La précoordination est recommandée quand un modificateur apparaît très souvent associé à un autre terme particulier.

3.2.7 ABREVIATIONS ET SIGLES

Il faut éviter, en général, les formes abrégées des termes, car leur emploi risque de ne pas être assez général, leur signification pouvant dépendre du contexte. En outre, leur compréhension peut dépendre de l'emploi de lettres capitales et de points séparateurs qui constituent des sujétions si des imprimantes de calculateurs ou d'autres matériels électroniques de traitement de l'information sont utilisés en liaison avec le thesaurus. Il faut donc limiter leur emploi aux cas où leur signification est bien assise, soit au sein du groupe d'utilisateurs intéressés, soit internationalement, et quand il est prouvé qu'ils assurent une sérieuse économie dans les applications.

La forme abrégée et la forme développée d'un terme donné doivent être considérées comme des synonymes et, en conséquence, faire l'objet des renvois voulus.

Les abréviations à plusieurs significations doivent être considérées comme des homonymes (homographes).

La nécessité de limiter la longueur des descripteurs oblige parfois à employer des abréviations moins courantes. Il faut alors les faire suivre d'une note d'application. Certains sigles très connus sont acceptés comme descripteurs.

Exemple : RADAR, LASER

3.2.8 EMPLOI DES CARACTERES

a) Jeux de caractères

L'emploi éventuel de matériels de traitement de l'information peut exiger :

- l'emploi des seuls caractères de haut de casse pour les descripteurs ;
- l'exclusion des signes diacritiques ;
- la limitation du nombre de caractères d'un descripteur.

Ces restrictions disparaissent quand les matériels de traitement de l'information sont adaptés aux besoins de l'information.

b) Ponctuation

L'emploi des signes de ponctuation, dans les descripteurs, doit être réduit. Sauf pour les nomenclatures spécialisées, seuls les parenthèses et les traits d'union sont nécessaires dans les descripteurs suivant les règles énoncées ci-dessous.

Le point n'est autorisé que lorsqu'un mot doit être tronqué parce que la longueur du descripteur est limitée. Il ne faut user de traits

d'union que lorsque leur omission modifierait le sens que l'on veut donner au descripteur. Il faut proscrire les virgules, les deux-points et les apostrophes quand le sens ne les exige pas. La virgule est autorisée en cas d'inversion. Là où les signes de ponctuation sont exclus, il est conseillé de s'en servir dans le libellé des notes d'application.

c) Caractères spéciaux et chiffres

Lorsqu'on juge nécessaire d'employer, dans les descripteurs, des caractères spéciaux autres que les traits d'union et les parenthèses, leur signification doit être clairement définie. D'autres caractères spéciaux peuvent être utilisés dans les notes d'application, définitions et autres formes d'information complémentaires. Si le descripteur contient des éléments numériques, seuls les chiffres arabes doivent être employés. Il convient de suivre les usages relatifs à la position des chiffres. Il faut aussi fixer des règles pour le traitement des signes diacritiques et des chiffres placés en indice supérieur ou en indice inférieur.

3.3 COMMENT EVITER L'AMBIGUÏTE

3.3.1 HOMONYMES (HOMOGRAPHES)

Les différents sens des homonymes (homographes) doivent être mis en évidence et distingués les uns des autres par des symboles ou des termes qualificatifs, faisant partie du descripteur et placés entre parenthèses immédiatement après l'homonyme. Ces symboles ou termes précisent le sens s'ils ne sont pas eux-mêmes des homonymes ou des signes utilisables ailleurs. L'homonyme et le qualificatif entre parenthèses forment ensemble un descripteur composé.

Exemple : CALCUL (INTEGRAL)
CALCUL (RENAL)

Lorsqu'un terme défini dans un vocabulaire technique normalisé international ou national est choisi comme descripteur, il faut l'enregistrer, sans aucune modification notable, y compris les parties du terme entre parenthèses dans l'article correspondant du vocabulaire.

3.3.2 NOTES D'APPLICATION (OU NOTES D'USAGE)

Une note d'application explique brièvement l'emploi auquel est destiné le descripteur. Ces notes peuvent accompagner le descripteur dans la partie principale du thesaurus mais ne font pas partie du descripteur.

Elles peuvent servir :

- à limiter l'emploi d'un descripteur ;
- à expliquer les abréviations et sigles ;
- à exclure un sens possible d'un terme, spécialement s'il s'agit de termes utilisés couramment dans différentes disciplines ;
- à indiquer les dates d'addition ou de suppression de termes et à signaler les changements dans le mode d'emploi des termes.

Les notes d'application doivent être composées en caractères spéciaux et distinguées clairement des qualificatifs.

Exemples : COUCHE MINCE (chimie analytique)
COUCHE PRIMAIRE (peinture)
PH (paraboloïde hyperbolique)

3.3.3 DEFINITIONS

Le contenu sémantique d'un descripteur est indiqué principalement par ses relations avec d'autres termes du thesaurus. Chaque fois qu'il subsiste un doute sur l'interprétation univoque d'un descripteur, il faut ajouter, pour en préciser le contenu sémantique exact, une définition qui peut accompagner le descripteur dans la partie principal du thesaurus mais ne fait pas partie du descripteur.

Exemple : DOCUMENTATION
(Processus de mise en mémoire et de recherche de l'information dans tous les domaines de la connaissance)

3.3.4 TRADUCTIONS

Il est souvent utile de donner les termes équivalents dans d'autres langues, pour s'assurer de l'emploi correct du descripteur dans l'analyse de textes en langue étrangère. Lorsque les sens ne sont pas entièrement équivalents on peut attirer l'attention sur ce point par une note explicative. Les traductions des descripteurs simples doivent être traitées comme des synonymes ou quasi-synonymes/1.

3.3.5 INDICATION DES SOURCES

L'indication de la source d'un descripteur ou d'une définition peut être très importante pour le développement futur du thesaurus. C'est pourquoi, quand on enregistre un descripteur, il peut y avoir lieu de noter en même temps la source, mais il n'est pas indispensable de faire figurer cette dernière dans la partie principale du thesaurus/2.

---

1. Le problème de la traduction de thesaurus fera l'objet de règles spéciales (en préparation).
2. Code 3166 de l'ISO pour la représentation des noms de pays (en préparation).

3.4 RELATIONS ENTRE LES DESCRIPTEURS

3.4.1 GENERALITES

Par définition, une fonction primordiale d'un thesaurus est de représenter les relations entre notions par l'indication des rapports entre les mots utilisés pour les décrire. Le réseau des relations d'un descripteur avec les autres descripteurs fournit ainsi une sorte de définition en situant le descripteur dans un champ sémantique.

Actuellement, trois sortes de relations doivent retenir l'attention d'un constructeur de thesaurus : ce sont les relations d'équivalence, les relations hiérarchiques et les relations associatives. Dans les trois cas, la relation est réciproque, c'est-à-dire que lorsqu'un descripteur est en relation avec un autre, cette relation doit être indiquée pour les deux descripteurs.

3.4.2 RELATION D'EQUIVALENCE (RELATION PREFERENTIELLE)

Les termes considérés comme équivalents (similaires, ou de signification presque identique) peuvent être réunis dans des catégories d'équivalence, de telle sorte que les termes équivalents correspondent à une seule et même notion. Dans la recherche documentaire, tous les documents associés à la catégorie d'équivalence doivent être retrouvés, même si un seul des termes est utilisé comme descripteur.

Il faut distinguer :

- les synonymes, c'est-à-dire les termes qui ont le même sens ou presque dans une discipline particulière :

  Exemple : CHANGEMENT
  MODIFICATION

- les quasi-synonymes, c'est-à-dire les termes dont la signification peut être différente dans le vocabulaire utilisé et le domaine concerné, mais qui sont regardés comme synonymes pour les besoins du système de documentation considéré.

a) Renvoi "USE" /EM (employer)/

Dans les systèmes à termes préférentiels, le symbole "USE" (EM) est employé pour renvoyer d'un non-descripteur à un ou plusieurs descripteurs aux fins suivantes :

- Indiquer un synonyme à préférer.
  Exemple : COURBURE EM FLEXION

- Renvoyer d'un terme particulier à un terme plus général (quasi-synonyme) qui a été choisi pour représenter la notion particulière.

  Exemple : CIRE VEGETALE EM CIRE

- Indiquer une orthographe à préférer, ou bien développer ou expliquer une abréviation.

  Exemple : MESONS PI EM PIONS

- Recommander l'emploi de deux termes ou plus pour exprimer une notion (division sémantique).

  Exemple : FILM FERROMAGNETIQUE EM MATERIAU FERROMAGNETIQUE + FILM

- Exprimer des notions pouvant être considérées comme synonymes (quasi-synonymes) pour les besoins de l'indexation et de la recherche documentaire.

  Exemple : HEREDITE EM GENETIQUE

- Rapprocher différents éléments ou degrés d'une suite continue de variantes d'une même notion.

  Exemple : FLUIDITE EM VISCOSITE

- Renvoyer d'un article disposé selon un ordre non naturel au terme écrit dans l'ordre correct des mots.

  Exemple : FONCTIONS, TABLES EM TABLES DE FONCTIONS

- Indiquer la terminologie courante.

  Exemple : CAPACITE ELECTRIQUE EM CONDENSATEUR

- Remplacer les termes du jargon technique par les termes corrects.

  Exemple : TERRE EM MISE A LA TERRE

- Donner des traductions de descripteurs.

  Exemple : X-RAYS EM RAYONS X

b) Renvoi "USED FOR" /EP (Employé pour)/

Réciproquement, le symbole "USED FOR" (EP) est employé avec le terme préféré, comme renvoi inverse. Il accompagne le terme auquel renvoie le symbole "USE" (EM).

c) Renvoi "USED FOR COMBINATION" /EPC (employé pour combinaison)/

Dans les cas où un descripteur est utilisé en combinaison avec d'autres descripteurs pour exprimer une notion représentée par un non-descripteur, il est bon de l'indiquer par le renvoi EMPLOYE POUR COMBINAISON (voir 3.3.2).

Il accompagne chaque terme auquel renvoie le symbole EM :

Exemple : MATERIAU FERROMAGNETIQUE EPC FILM FERROMAGNETIQUE
FILM EPC FILM FERROMAGNETIQUE

3.4.3 RELATIONS HIERARCHIQUES

Les relations hiérarchiques expriment les rapports de supériorité et de subordination entre les notions. On les subdivise comme suit :

a) Relation générique

Dans cette relation, le terme générique (supérieur) représente une classe de notions dont fait toujours partie la notion représentée par le terme spécifique (subordonné). La notion spécifique diffère de la notion générique par au moins une caractéristique.

Exemple :

Terme générique : MOTEUR A COMBUSTION INTERNE

Termes spécifiques : MOTEUR A PISTON — MOTEUR A ESSENCE
MOTEUR ROTATIF — MOTEUR DIESEL

b) Relation tout-partie (relation partitive)

Dans cette relation, le terme supérieur représente un objet ou un concept (entité) dont fait partie l'objet ou le concept (partie) représenté par le terme subordonné. Les termes désignant les parties s'obtiennent en imaginant la décomposition en ses composants de l'entité représentée par le terme supérieur.

Exemple :

entité : MOTEUR A PISTON

parties : PISTONS
CYLINDRES
BIELLES
VILEBREQUIN
etc.

Les relations hiérarchiques peuvent être représentées, selon le cas, à l'aide d'un des procédés suivants :

- les relations génériques et les relations partitives sont différenciées et exposées séparément ;
- les relations génériques et les relations partitives ne sont pas différenciées et sont groupées sous les renvois hiérarchiques.

Dans les disciplines où les relations partitives n'ont pas d'importance pour la recherche documentaire hiérarchique, il est recommandé de ne représenter que la relation générique sous le renvoi hiérarchique ; dans ce cas, la relation partitive est traitée comme une relation associative.

Il faut éviter, dans un thesaurus, de représenter une relation partitive par un renvoi à la fois hiérarchique et associatif.

Dans la plupart des thesaurus, la relation hiérarchique est représentée par le renvoi "BROADER TERM" (BT) (terme générique - TG) désignant la notion générique et par le renvoi inverse "NARROWER TERM" (NT) (terme spécifique - TS) désignant la notion subordonnée. Lorsque les deux types de relations hiérarchiques doivent être différenciés, il y a lieu de choisir des symboles différents pour les relations génériques et les relations partitives. Dans ce cas, il est conseillé d'employer les renvois suivants : "BROADER TERM GENERIC" (BTG) (terme générique dans une relation générique - TGG) et "NARROWER TERM GENERIC" (NTG) (terme spécifique dans une relation générique - TSG) pour les relations génériques et "BROADER TERM PARTITIVE" (BTP) (terme générique partitif - TGP) et "NARROWER TERM PARTITIVE" (NTP) (terme spécifique partitif - TSP) pour les relations partitives.

3.4.4 RELATION ASSOCIATIVE (relation d'affinité)

La relation associative est utilisée généralement pour recouvrir les autres relations entre les notions liées, mais en excluant les relations hiérarchiques et les relations d'équivalence (c'est-à-dire similarité, antonymie).

Il faut remarquer cependant qu'il existe divers types de relations entre notions. Il ne faut donc établir de relations associatives que lorsqu'il apparaît que ces relations sont nécessaires à la recherche documentaire. On peut signaler les notions associées par le renvoi "RELATED TERM" (RT) (Voir aussi - VA).

Les relations associatives peuvent être employées pour indiquer, par exemple :

- l'antonymie, lorsqu'une notion est l'opposé d'une autre notion : IMPERMEABILITE VA PERMEABILITE ;
- la coordination, lorsque des notions sont dérivées d'un terme générique au même niveau de classement : RELATION GENERIQUE VA RELATION PARTITIVE ;
- le rapport génétique, lorsqu'un objet est le prédécesseur d'un autre objet : PERE VA FILS ;

- l'emploi de deux notions coexistantes : EDUCATION VA ENSEIGNEMENT ;
- la cause et l'effet : ENSEIGNEMENT VA INSTRUCTION ;
- un rapport instrumental : ECRITURE VA CRAYON ;
- un rapport matériel, lorsqu'un objet est le matériau dont est fait un autre objet : PAPIER VA LIVRE ;
- diverses sortes de similarité (physique, matérielle, similarité de processus, etc.) : ENSEIGNEMENT VA FORMATION.

Cette liste a seulement pour objet de présenter des exemples des diverses sortes de rapports que la relation associative peut embrasser. Seules doivent être établies les relations utiles à l'indexation ou à la recherche documentaire.

Dans certains domaines, il peut être intéressant de représenter d'une façon explicite certaines de ces relations, en plus des relations hiérarchiques et associatives. Ces relations supplémentaires doivent être clairement définies et codées.

3.4.5 SYMBOLES DE RENVOI

Les relations peuvent être indiquées de différentes manières. Si l'on se sert, pour cela, d'un code, il faut toujours en préciser le sens.

3.4.5.1 Relations hiérarchiques

a) Lorsque les relations génériques et partitives ne doivent pas être différenciées ou que seule la relation générique est traitée comme une relation hiérarchique, il est recommandé d'utiliser les symboles suivants :

"BT" - TG TERME GENERIQUE (lorsque les relations génériques et partitives ne doivent pas être différenciées)

"NT" - TS TERME SPECIFIQUE (lorsque les relations génériques et partitives ne doivent pas être différenciées)

b) Lorsque les relations génériques et partitives doivent être différenciées, il est recommandé d'utiliser les symboles suivants :

au lieu de TG :

"BTG" - TGG TERME GENERIQUE dans une RELATION GENERIQUE
"BTP" - TGP TERME GENERIQUE PARTITIF (entité)

au lieu de TS :

"NTG" - TSG TERME SPECIFIQUE dans une RELATION GENERIQUE
"NTP" - TSP TERME SPECIFIQUE PARTITIF (partie).

3.4.5.2 Autres relations

Pour les autres relations, il est recommandé d'utiliser les symboles suivants :

"RT" - VA VOIR AUSSI (relation associative)

"USE" - EM EMPLOYER un synonyme, un quasi-synonyme ou une combinaison de descripteurs

"UF" - EP EMPLOYE POUR un synonyme, un quasi-synonyme ou une combinaison de descripteurs

"UFC" - EPC EMPLOYE POUR COMBINAISON, dans le cas de descripteurs simples.

Si un utilisateur ne peut adopter cette recommandation, il ne doit employer qu'une seule série de symboles compatibles avec ceux qui sont recommandés ci-dessus.

4. PRESENTATION ET DISPOSITION DU THESAURUS

4.1 INTRODUCTION

Tout thesaurus doit être précédé d'une introduction détaillée, qui indique clairement son objet, sa structure et les domaines qu'il couvre. Il faut signaler les règles qui ont présidé à son élaboration, ainsi que les méthodes et les services auxquels on a fait appel pour choisir les termes, fixer leur forme et éviter toute ambiguïté. Il convient aussi d'indiquer avec précision le mode de présentation, la signification des abréviations employées et, chaque fois qu'il y a lieu, les règles suivies pour le classement alphabétique et la ponctuation.

Il faut enfin, et c'est le plus important, exposer le mode d'emploi du thesaurus et dire quelles sont ses limites en donnant des exemples le cas échéant.

Les utilisateurs doivent être invités à formuler des observations et des suggestions en vue d'aménagements éventuels, et à faire savoir s'ils veulent recevoir les éditions ultérieures ou les suppléments qui seront publiés. Il faut exposer selon quelles modalités on envisage de développer le thesaurus et de le mettre à jour ; il faut indiquer la date de parution et la date probable des éditions ultérieures, ou de la publication de suppléments.

Il convient de préciser le nombre des descripteurs et des non-descripteurs.

4.2 PARTIE PRINCIPALE DU THESAURUS

La partie principale doit fournir une information complète sur chaque descripteur (dans les systèmes à termes préférentiels) ou sur chaque notion (dans les systèmes n'employant pas de termes préférentiels). On peut présenter la partie principale de façon systématique, ou de façon alphabétique, ou combiner

les deux présentations. Lorsque la partie principale est alphabétique et comprend des non-descripteurs et des permutations de descripteurs composés et de non-descripteurs, un index alphabétique séparé n'est pas nécessaire. Un thesaurus doit toujours comprendre une présentation systématique et une présentation alphabétique.

Les informations suivantes doivent figurer dans la partie principale :

Représentations de notions

Le code numérique de la notion et/ou sa ou ses notation(s)
Le ou les descripteur(s)
Le ou les non-descripteur(s)
Les renvois "EMPLOYE POUR" (relations d'équivalence) ; les renvois "EMPLOYE POUR COMBINAISON" dans le cas de descripteurs simples.

Informations complémentaires

définition(s)
notes d'application
brève information sur la source

Relations entre notions

notions supérieures (relations hiérarchiques) - TG
  notions génériques (relations génériques) - TGG
  notions d'entité (relations partitives) - TGP

notions subordonnées (relations hiérarchiques) - TS
  notions spécifiques (relations génériques) - TSG
  notions de partie (relations partitives) - TSP

autres relations particulières
  notions associées (relations associatives) - VA

Il faut remarquer cependant que des thesaurus sont polyhiérarchiques et/ou ont une structure à facettes, si bien que plusieurs notions génériques ou spécifiques peuvent correspondre à un terme donné. Comme on ne peut présenter avec cohérence, dans un classement systématique, qu'une seule chaîne hiérarchique, les relations avec les autres chaînes hiérarchiques doivent donc être précisées par des renvois.

4.3 PARTIES ANNEXES DU THESAURUS

4.3.1 Généralités

Pour rendre plus accessible la partie principale du thesaurus, celui-ci peut être complété par plusieurs parties annexes telles que : index alphabétique, liste systématique et représentation graphique.

Un index alphabétique est nécessaire lorsque :

- la partie principale est ordonnée de façon systématique ;
- la partie principale est ordonnée de façon systématique et alphabétique à la fois ;
- la partie principale n'enregistre que les descripteurs (dans le cas de systèmes à termes préférentiels).

Une liste systématique est nécessaire lorsque :

- la partie principale est ordonnée alphabétiquement ;
- la partie principale est ordonnée de façon systématique et alphabétique à la fois.

Les parties annexes renvoient à l'article approprié de la partie principale. Pour cela, il est utile d'attribuer aux descripteurs ou aux notions, des codes numériques ou une notation équivalente.

4.3.2 Index alphabétiques

Les index doivent comprendre les descripteurs et les non-descripteurs.

Il est souhaitable de permuter les descripteurs composés et les expressions composées.

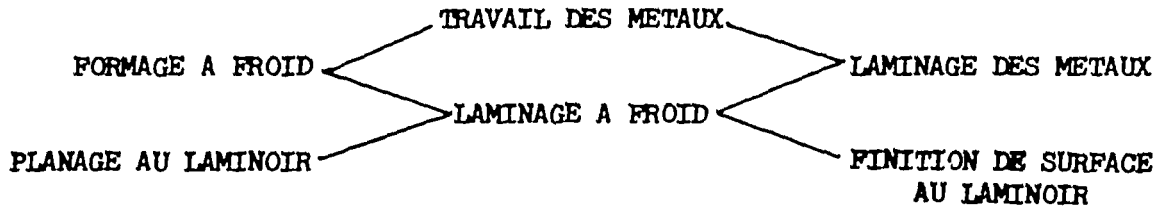
Les index alphabétiques peuvent se présenter sous la forme d'un index permuté ou d'un index de mots clés dans leur contexte (index KWIC). Il est préférable d'utiliser les normes nationales et les recommandations internationales pour la disposition alphabétique. Comme il existe des méthodes différentes suivant les domaines, il importe, avant de commencer le classement, de préciser les règles de mise en ordre alphabétique.

4.3.3 Listes systématiques

Les listes systématiques doivent comprendre tous les descripteurs (dans les systèmes à termes préférentiels) ou toutes les notions (dans les systèmes n'utilisant pas de termes préférentiels), conformément aux relations hiérarchiques représentées dans le thesaurus. Si, dans le thesaurus, on emploie plus d'une sorte de relation hiérarchique, les relations peuvent être combinées dans une seule liste ou précisées dans des listes séparées. Les listes systématiques doivent mettre en évidence les différents niveaux hiérarchiques. Si le thesaurus est polyhiérarchique, il est nécessaire de dresser la liste des termes communs à plusieurs chaînes hiérarchiques pour la faire figurer chaque fois à sa place dans la représentation hiérarchique. Si un terme est relié hiérarchiquement à plus d'un terme générique au niveau immédiatement supérieur, il est préférable de faire

également la liste des autres termes génériques qui s'y rapportent, en marquant le niveau hiérarchique par un symbole approprié (par exemple ♦) et par un décalage à droite.

Exemple :



Dans une liste systématique, cette structure peut être représentée de la façon suivante ("-" = spécifique "♦" = générique) :

TRAVAIL DES METAUX
- FORMAGE A FROID
- - LAMINAGE A FROID
♦ LAMINAGE DES METAUX
♦ ♦ TRAVAIL DES METAUX
- - - PLANAGE AU LAMINOIR
- - - FINITION DE SURFACE AU LAMINOIR
- LAMINAGE DES METAUX
- LAMINAGE A FROID
♦ FORMAGE A FROID
♦ ♦ TRAVAIL DES METAUX
- - - PLANAGE AU LAMINOIR
- - - FINITION DE SURFACE AU LAMINOIR

Cette présentation met en évidence les structures polyhiérarchiques.

La présentation sous forme de liste systématique convient probablement mieux aux domaines très spécialisés qu'aux domaines interdisciplinaires.

La combinaison de cette disposition avec un index alphabétique donne une liste alphabétique structurée, qui est probablement celle qui permet d'exploiter au maximum les avantages des deux présentations.

4.3.4 Représentations graphiques

Présenter les descripteurs et les relations entre eux sous la forme d'un schéma constitue peut-être le mode de présentation le plus nuancé que l'on puisse donner à un thesaurus. On peut concevoir un schéma à plusieurs dimensions, par exemple en donnant deux dimensions à chaque facette d'un thesaurus qui en comporte plusieurs, mais le schéma le plus usité actuellement est à deux dimensions.

Un tel système consiste à répartir les descripteurs en groupes sémantiques, à disposer chaque groupe sur une grille, et à assigner à chaque descripteur une place déterminée par rapport aux deux axes (horizontal et vertical) de telle sorte qu'il est doté de coordonnées.

On peut mettre en évidence les relations entre descripteurs, dans une représentation graphique, par :

- une disposition graphique (formation de groupes, inscription dans un système de coordonnées, utilisation de niveaux d'abstraction différents, cercles concentriques symboliques indiquant les différents niveaux hiérarchiques).
- des flèches entre les descripteurs (flèches à double sens pour les relations associatives indiquant le descripteur le plus spécifique, parenthèses plus une flèche qui part du terme préférentiel ou qui y aboutit pour les relations d'équivalence).

Il est entendu qu'un descripteur peut appartenir à plusieurs groupes. La dimension optimale de chaque groupe se situe, semble-t-il, entre trente et quarante termes. Comme pour la présentation antérieure, il faut donner en annexe une liste alphabétique indiquant le(s) groupe(s) sémantique(s) dont chaque descripteur relève. On choisira le mode de présentation le plus compatible avec l'usage auquel le thesaurus est destiné.

4.4 ORDRE DE CLASSEMENT (CLASSEMENT ALPHABETIQUE)

Pour les listes alphabétiques de la partie principale du thesaurus ou pour l'index alphabétique, il faut définir l'ordre de classement des lettres, des chiffres, des signes de ponctuation et des symboles particuliers.

Le classement peut se faire :

a) <u>lettre par lettre</u>, en suivant les règles ci-après :
   - ne pas tenir compte des espaces entre les mots ;
   - ne pas tenir compte des caractères autres que les parenthèses de gauche, les chiffres et les lettres ;
   - aligner dans cet ordre : parenthèse de gauche, chiffres en ordre croissant, lettres de a à z.

b) <u>mot par mot</u>, en suivant les règles ci-après
   - chaque mot complet est considéré à tour de rôle et classé lettre par lettre ;
   - tous les termes commençant par un mot complet donné sont placés avant tout terme commençant par la même suite de lettres qui constitue une partie du mot ;
   - les caractères non alphanumériques (signes de ponctuation ou caractères particuliers) sont considérés comme des espaces.

c) par tri sur machine, en suivant les règles ci-après :

- tous les caractères à l'intérieur d'un terme, y compris les lettres, les signes numériques, les signes de ponctuation et les caractères particuliers entrent en ligne de compte dans le classement ;
- le nombre d'espaces à l'intérieur d'un terme doit être fixé ;
- le classement s'effectue en suivant l'ordre des chiffres considérés individuellement, et non la valeur numérique ;
- le classement exact doit être basé sur la configuration du calculateur utilisé.

On adopte un mode de classement alphabétique en fonction de tous les éléments qui entrent en ligne de compte dans l'élaboration du thesaurus, c'est-à-dire de l'étendue et de la structure des domaines recensés, de la possibilité de recourir au traitement par ordinateur, du type de matériel disponible, etc. Dans tous les cas, il faut fixer clairement et explicitement les modalités du classement alphabétique, avant tout essai d'ordonnancement.

5. METHODE DE CONSTRUCTION DU THESAURUS

5.1 COMMENT EVITER LES DOUBLES EMPLOIS

Avant d'entreprendre l'établissement d'un thesaurus, il convient de s'assurer qu'il n'en existe pas d'autres qui traitent du même domaine ou d'un domaine voisin. A cette fin, il est conseillé de se renseigner auprès des deux centres d'information mentionnés ci-dessous[1]. Même si un ouvrage analogue est habituellement utilisé, on ne renoncera peut-être pas à établir celui que l'on a prévu. En pareil cas, il y a lieu d'indiquer clairement, dans l'introduction, les raisons de cette décision et les différences entre le nouveau thesaurus et l'ancien.

5.2 NOTIFICATION PREALABLE

Le centre d'information compétent (voir ci-dessus) doit être avisé de l'intention d'établir un thesaurus et être informé de la date de la première publication et diffusion. Cette notification doit être transmise par l'intermédiaire de l'organisme national qui s'occupe des thesaurus, lorsqu'il en existe un.

---

1. Pour les thesaurus de langue anglaise :

   Bibliographic Systems Center, School for Library Science, Case, Western Reserve University, Cleveland, Ohio 44106, USA.

   Pour les thesaurus établis dans d'autres langues :

   Centralny Institut Informacji Naukowo Technicznej i Ekonomicznej (CIINTE) - Clearinghouse Al, Niepodległości 188, Varsovie, Pologne.

La même recommandation vaut pour les éditions ultérieures. Autant que possible, il faut envoyer au centre d'information en question un exemplaire du thesaurus accompagné de son introduction et de son index. Le fait que le centre a été avisé doit être mentionné dans l'introduction.

5.3 PRINCIPES GENERAUX

Avant de décider de la structure générale du thesaurus, il faut étudier la structure du vocabulaire spécialisé considéré ainsi que les besoins des utilisateurs du système d'information et de documentation.

La sélection des descripteurs ne doit commencer qu'une fois décidée la structure générale du thesaurus. Il est recommandé :

- de confier à quelques personnes employées à plein temps les parties les plus longues du travail et la coordination des travaux effectués par les spécialistes sur chaque sujet ;
- d'avoir le plus possible d'experts ayant une bonne connaissance du domaine à traiter ainsi qu'une expérience de l'indexation et de la recherche ;
- de former des groupes de travail pour chaque secteur spécialisé du domaine à couvrir ;
- de faire appel à des groupes de travail internationaux, afin d'élargir l'expérience linguistique nécessaire à la construction du thesaurus.

La méthode de sélection des descripteurs varie selon la structure choisie pour le thesaurus (index alphabétique, liste systématique, représentation graphique), la destination du thesaurus (par exemple, pour la recherche documentaire, manuelle ou mécanique, seulement pour l'indexation ou comme outil secondaire) et l'arrière-plan du projet (construction progressive en vue d'un traitement mécanique, introduction d'un nouveau domaine, par exemple zones interdisciplinaires, pour lesquelles il n'existait encore aucun plan de classification; existence de groupes bien définis d'utilisateurs et de spécialistes ; documentation étendue).

5.4 DIFFERENTES METHODES

En général, il y a deux méthodes pour construire un thesaurus. Le choix dépend du type et du domaine de l'ensemble de documents que le thesaurus visera à couvrir.

Ces deux méthodes sont la méthode analytique et la méthode synthétique. Toutes deux sont empiriques, en ce sens que ni l'une ni l'autre ne reposent sur une analyse linguistique rigoureuse. En général, la méthode analytique, qui comprend l'analyse du contenu de l'ensemble des documents et le choix des termes à

partir de celle-ci, est préférable, surtout dans les domaines spécialisés de la connaissance. La méthode synthétique fait appel à des experts qui analysent la terminologie fournie par des sources secondaires telles que les index et autres documents de référence. Ces experts choisissent les termes préférentiels et déterminent les relations entre les termes. Cette méthode s'applique plus généralement aux domaines étendus comprenant plusieurs disciplines.

Dans la pratique, on utilise souvent une combinaison des deux méthodes. Selon la méthode analytique, on commence par indexer une série de documents, dans les domaines de la connaissance que le thesaurus vise à couvrir.

A partir des données obtenues par cette opération d'indexation expérimentale, le constructeur du thesaurus essaye d'établir des relations valables pour le système de recherche documentaire considéré, puis corrige le thesaurus grâce à une nouvelle indexation de la série de documents initiale et en étendant l'indexation à de nouveaux documents.

La méthode synthétique a besoin de partir d'une vue d'ensemble de la structure de la terminologie dans le domaine étudié, la structure étant en effet une taxonomie ou une classification. On demande à des experts connaissant bien le sujet de reconnaître ces relations ou structures en examinant les termes expérimentalement choisis pour décrire les documents. Au contraire de la méthode analytique, le thesaurus n'est pas construit par l'expérience pratique mais par des contributions faisant déjà autorité.

5.5 CHOIX DES DESCRIPTEURS

En général, le choix des descripteurs se fait en quatre étapes distinctes : collecte de termes, vérification, évaluation et choix proprement dit.

5.5.1 Collecte des termes

Il est pratiquement impossible de réunir tous les termes susceptibles de constituer des descripteurs si l'on pense à une liste alphabétique. Lorsqu'on envisage les descripteurs par groupes, les associations d'idées entre ceux-ci font apparaître de nombreux candidats.

Les sources pour ces collectes de termes[1] peuvent être :

- les utilisateurs éventuels et les spécialistes des sujets traités ;
- les dictionnaires techniques normalisés, nationaux ou internationaux ;
- la documentation actuelle ;

---

1. Il est possible de consulter le Centre international d'information pour la terminologie dont l'adresse est la même que celle de l'Osterreichisches Normungsinstitut (Institut autrichien de normalisation), Vienne (Autriche).

- les traités ou les listes terminologiques ;
- les thesaurus existants et les plans de classification systématiques ;
- les nomenclatures ;
- les index des revues techniques ;
- les services d'analyse ;
- les ouvrages d'enseignement, les manuels et les résumés ;
- les indexations expérimentales de documents.

Il existe dans certains domaines des nomenclatures extrêmement spécialisées ou bien des vocabulaires techniques normalisés qui font autorité. Toutes les fois qu'elle existe, il ne faut pas manquer de faire appel à une nomenclature déjà admise sur le plan international. Si l'on multiplie les dénominations particulières sans lien entre elles, on risque de faire du thesaurus une simple liste d'identificateurs, ce qui ôterait à l'ouvrage toute son utilité. Il est donc recommandé d'éviter, autant que possible, les noms d'entités particulières sans lien entre elles. Il est possible d'établir, indépendamment de la partie principale du thesaurus, une liste des noms spécifiques qui ne sont pas considérés comme des descripteurs.

5.5.2 Vérification

Quelle que soit la méthode de collecte adoptée, il faut vérifier la valeur scientifique des descripteurs choisis, en consultant des dictionnaires, d'autres vocabulaires d'indexation ou vocabulaires normalisés, en se référant à l'usage, et surtout en demandant leur avis à des spécialistes du sujet. Il faut écarter les termes qui ne sont plus usités ou les retenir uniquement pour signaler qu'ils sont proscrits.

L'une des qualités les plus appréciables d'un thesaurus est sa capacité d'assimiler immédiatement les néologismes et le jargon spécialisé qui se multiplient dans les domaines en expansion de la recherche fondamentale et de la recherche appliquée. Il faut tirer pleinement parti de cette capacité et recourir, en même temps, à des notes d'application et à des renvois. Il convient d'accorder une attention particulière aux termes dont le sens a évolué avec le temps, ou dont la signification varie d'un pays à l'autre. Au cas où il faut retenir des termes qui se recoupent, il convient de le signaler par le renvoi voulu.

Exemple : billion
(10^9 aux Etats-Unis)
billion
(10^{12} en France)

5.5.3 Evaluation

Pour évaluer l'utilité des descripteurs possibles, il faut rechercher, pour chacun d'eux :

a) s'il apparaît fréquemment dans la littérature technique ou dans les informations existantes ;

b) quelle sera sa fréquence probable dans les demandes formulées ;

c) quelle est sa relation avec les descripteurs déjà acceptés ;

d) s'il est bien à sa place dans le vocabulaire actuel de la discipline considérée, et si sa valeur est réelle ;

e) s'il connote et dénote bien et commodément la notion en question.

Aucun de ces facteurs d'appréciation ne doit être envisagé isolément et il faut prêter une attention toute particulière aux domaines périphériques, pour lesquels on ne demande pas au descripteur d'être aussi parfaitement exhaustif et spécifique qu'en ce qui concerne le domaine principal.

5.5.4 Choix

Dans tous les cas, les descripteurs doivent être choisis aux fins d'insertion dans le thesaurus en fonction de leur efficacité probable du point de vue de la recherche documentaire et de leur importance dans les documents à indexer.

5.6 PROCEDE D'ENREGISTREMENT

Les procédés d'enregistrement des informations, au cours de la construction du thesaurus, peuvent varier en fonction des besoins spécifiques de chaque système (voir 5.3). C'est pourquoi seules quelques indications pratiques peuvent être données.

Le processus de sélection des termes et l'établissement de relations entre eux rendent nécessaire l'enregistrement sous une forme organisée de l'information recueillie sur un terme spécifique. Il faut donc mettre au point et utiliser, au cours de ce processus, un format spécifiant les éléments d'information qui entrent dans la construction du thesaurus (voir 4.2). L'aspect physique du format dépend du système que l'on veut réaliser et du matériel disponible. Ce peut être un imprimé, une fiche à perforation manuelle, une carte perforée ou un jeu de catégories de longueur variable, utilisables pour l'entrée sur ordinateur. Chaque fois que l'on considère un terme comme descripteur ou non-descripteur possible, il faut l'enregistrer. Les relations entre notions, contenues dans les sources d'où est extrait le terme candidat, doivent être enregistrées au cours de cette sélection. Pour chaque cas, il faut vérifier si les relations données par les sources sont fonction du contexte ou sont valables pour tous les cas.

Pendant l'élaboration du thesaurus, il est possible de prendre des décisions modifiables ultérieurement ; ce qui assure une certaine souplesse, mais exige de la part du constructeur des notations appropriées pour chaque terme affecté par une décision. Par exemple, pour la construction de "terme A employer le terme B", il faut faire figurer la note "USE" (EM) après le terme A et la note réciproque "USED FOR" (EP) après le terme B. De même pour les autres types de renvois. De cette façon, on peut modifier une décision à tout moment ; la forme de la note fournit les informations nécessaires pour garder au thesaurus sa cohérence interne. Cela permet de repérer tous les articles qui doivent être modifiés à la suite de la nouvelle décision.

Lors du choix des descripteurs, les termes choisis doivent être groupés systématiquement, par exemple par points de vue. Les notions représentées par ces descripteurs doivent être éclaircies, et il convient de comparer chaque descripteur aux autres en ce qui concerne les relations entre notions (selon la structure proposée du thesaurus). Il faut consulter les spécialistes et les utilisateurs éventuels sur le choix des descripteurs et sur les relations entre notions.

5.7 ESSAIS D'APPLICATION

Avant de donner au thesaurus ses bases définitives, il est très fortement recommandé de faire des essais pratiques, fondés sur un nombre suffisant de documents se rapportant au domaine à couvrir.

Ces essais doivent montrer :

- si la structure proposée du thesaurus tient compte des besoins les plus récents ;
- si les descripteurs choisis sont utiles pour l'indexation et la recherche ;
- dans quels domaines d'autres descripteurs sont nécessaires ;
- si les relations établies entre les notions sont correctes et suffisantes.

Il convient de demander des commentaires critiques sur ces essais au plus grand nombre possible de personnes, notamment à des spécialistes de l'information et de l'indexation, à des spécialistes du sujet, ainsi qu'à des utilisateurs.

6. MISE A JOUR

Le thesaurus doit être mis à jour de façon continue, en tenant compte des questions posées en 5.7. Cette mise à jour, selon une méthode déterminée, doit être confiée à une autorité centrale désignée à cet effet.

6.1 VERIFICATION PERIODIQUE

Il faut contrôler périodiquement la fréquence d'usage des descripteurs, que ce soit pour l'indexation ou pour la recherche. Il faut s'assurer, par des vérifications périodiques, que certains descripteurs n'en chevauchent pas d'autres ou ne font pas double emploi, et que les relations établies à l'intérieur du thesaurus restent exactes et satisfaisantes. Chaque fois qu'une recherche ne permet pas de localiser l'information voulue, ou ne permet pas de recueillir toute l'information que l'on croit pouvoir recueillir, il faut procéder à une évaluation critique des descripteurs qui ont été employés ou qui auraient dû l'être.

Lorsque des descripteurs deviennent obsolètes du fait de modifications du domaine ou de la terminologie, il faut les éliminer ou les remplacer par des descripteurs plus appropriés.

6.2 ELIMINATION DES DESCRIPTEURS

Lorsque l'on constate que la fréquence d'utilisation est très faible, il faut s'assurer que ce n'est pas dû exclusivement au manque de documents se rapportant à cette notion particulière. Théoriquement, on ne devrait supprimer un descripteur particulier que s'il n'a jamais été utilisé, ni pour l'indexation, ni pour la recherche documentaire. Il est plus commode d'indiquer par une relation préférentielle (renvoi EMPLOYER) où se situe, le cas échéant, le terme de remplacement.

Un descripteur qui a été éliminé peut être conservé comme synonyme ou être accepté pour la recherche uniquement afin d'éviter la nécessité de réindexer les documents qui utilisaient ce descripteur. A chaque fois il faut examiner et réadapter les relations du descripteur éliminé ou modifié.

D'autre part, si l'on rattache trop d'éléments à un même descripteur, celui-ci perd de sa spécificité. Il devient d'une application trop générale et il faut envisager de le fractionner. La décision à prendre, lorsque la fréquence d'utilisation d'un descripteur donné est trop forte ou trop faible, dépend, dans une certaine mesure, de la méthode employée pour la recherche documentaire. Si le descripteur le moins spécifique est celui sur lequel on exécute la recherche en dernier lieu, cela ne vaut peut-être pas la peine de le supprimer.

6.3 CHOIX DE NOUVEAUX DESCRIPTEURS

Si, pendant l'indexation et/ou la recherche documentaire, on constate que des notions ou que des relations entre notions n'ont pas été suffisamment précisées dans le thesaurus, il faut trouver un nouveau descripteur ou de nouvelles relations.

De plus, les indexeurs et les utilisateurs doivent constamment vérifier s'il n'apparaît pas de termes susceptibles de constituer des descripteurs nouveaux, représentant soit des notions nouvelles, soit des points de vue différents relatifs à des notions anciennes ou employés comme synonymes de termes existants. Il faut évaluer l'utilité des termes nouveaux et des nouvelles relations avant de les admettre définitivement.

Si possible, le nouveau descripteur doit être mis à l'essai par des indexeurs durant un certain temps, avant de devenir une addition définitive au thesaurus.

La fréquence d'utilisation de ces descripteurs éventuels, pour l'indexation ou la recherche rétrospective, est une indication sûre de leur utilité future. Une fois qu'il est décidé d'ajouter au thesaurus un descripteur nouveau, il convient de déterminer les relations qui existent entre ce terme et tous les descripteurs inscrits avant lui, et les indiquer aux endroits voulus.

Il faut éviter d'introduire un à un des descripteurs nouveaux, ce qui est une source de confusion pour l'utilisateur. Il vaut mieux les accumuler pendant un certain temps et les introduire ensemble, soit sous forme de "supplément au thesaurus" ou à l'occasion d'une nouvelle édition de l'ouvrage. Cela ne doit pas empêcher les indexeurs de les employer. Il devrait exister un service central chargé d'étudier toutes les suggestions qui lui parviennent et de se prononcer sur la recevabilité des additions éventuelles.

Il ne faut jamais oublier que le temps seul décide de la taille et de la forme d'un thesaurus : l'ouvrage n'est jamais terminé.

7 REFERENCES

7.1 LISTE DES RECOMMANDATIONS DE L'ISO LIEES AUX PRINCIPES DIRECTEURS CI-DESSUS

ISO/R 9	- 1968	Système international pour la translittération des caractères slaviques cyrilliques
ISO/R 233	- 1961	Système international pour la translittération des caractères arabes
ISO/R 259	- 1962	Translittération de l'hébreu
ISO/R 639	- 1967	Indicatifs de langue, de pays et d'autorité
ISO/R 704	- 1968	Principes de dénomination
ISO/R 843	- 1968	Système international pour la translittération des caractères grecs en caractères latins
ISO/R 860	- 1968	Unification internationale des notions et des termes
ISO/R 919	- 1969	Guide pour l'élaboration des vocabulaires systématiques (Exemple de méthode)
ISO/R 1087	- 1969	Vocabulaire de la terminologie
ISO/R 1149	- 1969	Présentation des vocabulaires systématiques multilingues

On peut obtenir les documents ci-dessus auprès des organisations nationales de normalisation correspondantes dans les pays membres de l'ISO.

7.2 SOURCES POUR LES DICTIONNAIRES, GLOSSAIRES ET THESAURUS

Bibliographie de dictionnaires scientifiques et techniques multilingues, 5e édition, Paris, Unesco, 1969, 250 p.

Bibliographie des vocabulaires scientifiques et techniques monolingues, Vol. I : Normes nationales, 1955, 210 p., Vol. II : sources diverses, 1959, 146 p., Paris, Unesco
(Des suppléments sont publiés dans Babel, Revue internationale de la traduction publiée par la Fédération internationale des traducteurs avec le concours de l'Unesco. Avignon, France).

Bibliographic Bulletin of the Clearinghouse at CIINTE, 1969, Varsovie, CIINTE, 1969, 140 p. (On compte publier des suppléments annuels)

Bibliographic Systems Center Subject Index, Case Western Reserve University, Cleveland, Ohio, (Etats-Unis d'Amérique), 1969.

(Index produit par imprimante).